Introduction to Big Data

Lesson 1.5 Social-issues

Arthur Katossky & Rémi Pépin

Ethical issues

- **Privacy** : more data are collected and analyzed, but in what extend and for what ?
- **Discrimination** : predictive justice, recruitment
- **Consent** : do individuals agree to give their data ? Do they really understand ? Have the choice ?
- **Transparency** : how models work ? What are their purpose ?
- **Security** : more data collected = more risk of data breaches = more cyber attacks
- **Captology** : using data to influence people or keeping then engage (*doomscrolling*)

Engaging all of these aspects is hard and clearly a team job. One may want to use reference guidelines. Take a look at the Deon project for an instance.

Are ML model discriminating ?

- Twitter investigates racial bias in image previews
- Amazon's sexist AI recruiting tool: how did it go so wrong?
- Predictive policing algorithms are racist. They need to be dismantled.
- Google Photos Tags Two African-Americans As Gorillas Through Facial Recognition Software

Do data scientists do it on purpose ?

NO ! (most of the time)

- ML model learn from input data
- Most data collection is biased towards well-off young males of European ancestry
- Social prejudices are embedded into data and retrieved by algorithms

Possible solutions :

1. have a team with multiple profiles with multiple points of view
2. always allow user feedbacks and do not dismiss criticism
3. plan ahead about the potential harms of you data-based solution

Environmental issues

- **Energy consumption**

  The Internet : 10% electricity consumption worldwide

  1 data center = consumption of 30 000 habs [1]

- **Matter matters**

  Computers use metals (including rare earths), usually mined in developing countries wit low regulation and releasing tremendous amounts of pollutants

- **Carbon Emissions / Water consumption**

  Building and running data centres is polluting.

  Data centre operator try to build green data centre by producing their own energy. Data centers can be couples with district heating (ex: Microsoft in Finland).

Big infrastructures are not always the most environmentally-friendly solution. Active work currently for measuring and mitigating environmental impacts of computing.

[1] Source

Political issues

- **Surveillance** : Easy to watch people in the current era. Like targeted add, or detecting odd behaviour with pattern detection.
- **War** : A lot of the AI research is directed towards autonomous weapons.
- **Misinformation** : Target contents to polarize people (electoral manipulation), distilled propaganda (Chinese video on TikTok). Not new phenomenon, big data helps to industrialize it.
- **Transparency** : Algorithmic decisions that impact people's life (grants, asylum, access to essential services, etc.) deserve to be explicit and transparent ; however this exposes the decision to adversarial selection
- **Sovereignty** : currently Europe is dependent of American companies. OVH growing but sill behind. We are financing US company by using their services.

Conclusion

It's a complex problem for ours democraties. Big data is here and will stay. Peoples need to be educated to understand the current state of the world.